

Using Wittgenstein's family resemblance principle to learn exemplars

Sunil Vadera^{*}, Andres Rodriguez⁺, Enrique Succar[^], J. Wu^{*}

^{*} University of Salford,

⁺ Instituto de Investigaciones Electricas, Mexico

[^] ITESM, Campus Morelos, Mexico.

ABSTRACT

The introduction of the notion of family resemblance represented a major shift in Wittgenstein's thoughts on the meaning of words, moving away from a belief that words were well defined, to a view that words denoted less well defined categories of meaning. This paper presents the use of the notion of family resemblance in the area of machine learning as an example of the benefits that can accrue from adopting the kind of paradigm shift taken by Wittgenstein.

The paper presents a model capable of learning exemplars using the principle of family resemblance and adopting Bayesian networks for a representation of exemplars. An empirical evaluation is presented on three data sets and shows promising results that suggest that previous assumptions about the way we categories needs reopening.

1. Introduction and Background

Much of our daily reasoning appears to be based on stereotypes, exemplars and anecdotes. Yet, basic statistics informs us that decisions based only limited data are, at best, likely to be inaccurate, if not badly wrong. However, exemplars and stereotypes are not arbitrary data points, they are based on experience and represent prototypical situations. The ability to predict the behaviour of a consumer, observe that two people are related, diagnose an illness, and even how an MP might vote on a particular issue, all depend on a person's past experience – that is the exemplars and stereotypes a person learns.

If this hypothesis, namely that we can form and reason well with exemplars, is true, we should be able to identify exemplars from data. To achieve this, we need to answer the following questions:

- (a) What is an exemplar and how can it be represented?
- (b) How do we learn good exemplars incrementally?
- (c) How can exemplars be used?

This paper presents a particular approach to these questions that involves the use of the notion of *family resemblance* to learn exemplars and Bayesian networks to represent and utilise exemplars. Section 2 of the paper presents the problem and our model that aims answers to the above questions. Section 3 presents the results of an empirical evaluation and section 4 presents the conclusion and future work.

2 The Problem and the Model

The central problem can be visualized as moving from a situation like that in figure 1 which has three categories A,B,C with a lot of data to one like that in figure 2, where we have exemplars representing the categories:

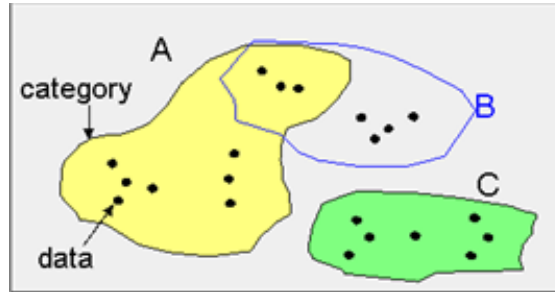


Figure 1: Categories with data

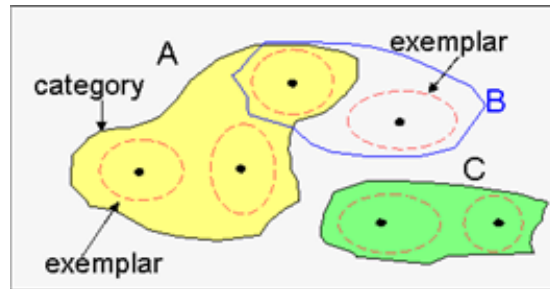


Figure 2: Exemplar based view

Given that membership of categories and the extent to which exemplars represents other points is graded, we use Bayesian networks (Pearl, 1991) to represent exemplars. Figure 3 depicts the Bayesian network representation used, where the e_i denote exemplars associated with the categories, the f_i denote the features and the arcs denote the dependencies. The node V_e is a virtual exemplar which is introduced to take account of all the data points that have not been seen as the model learns incrementally. It is needed to satisfy the conditions of the particular kind of network that we use, called the Noisy-Or model (See Pearl, 1991, for a description), that enables us to adopt a more efficient propagation algorithm. The dependencies $P(f_i|e_i)$ are estimated using a standard beta-distribution and $P(f_i|V_e)$ are estimated using a decay function as described in an earlier paper (Rodriguez and Vadera, 1999).

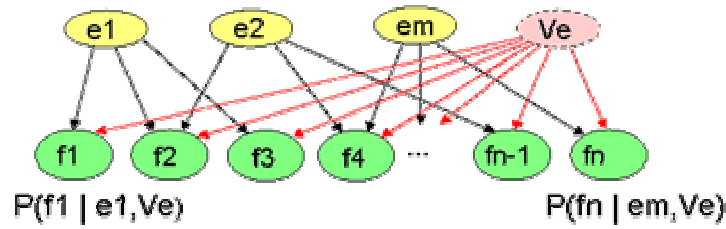


Figure 3: Bayesian network representation of an exemplar based model

Given such a model, assessing whether a new point is represented by an exemplar, is computed by using a propagation algorithm to compute the probability of each exemplar given the features of the new point. The exemplar with the highest probability can then be used to determine the category of the point.

But how do we learn such a model incrementally? As an initial experiment, a simple, greedy learning strategy, shown in figure 4, is adopted. The figure shows three exemplars, e1, e2, e3 and a new training case. An attempt is made to classify the case using the existing model, if this fails then it is added as a new exemplar (figure 4(a)). If it succeeds, then the exemplar that helps classify it and the new training point compete, with the better one being retained (figure 4(b)).

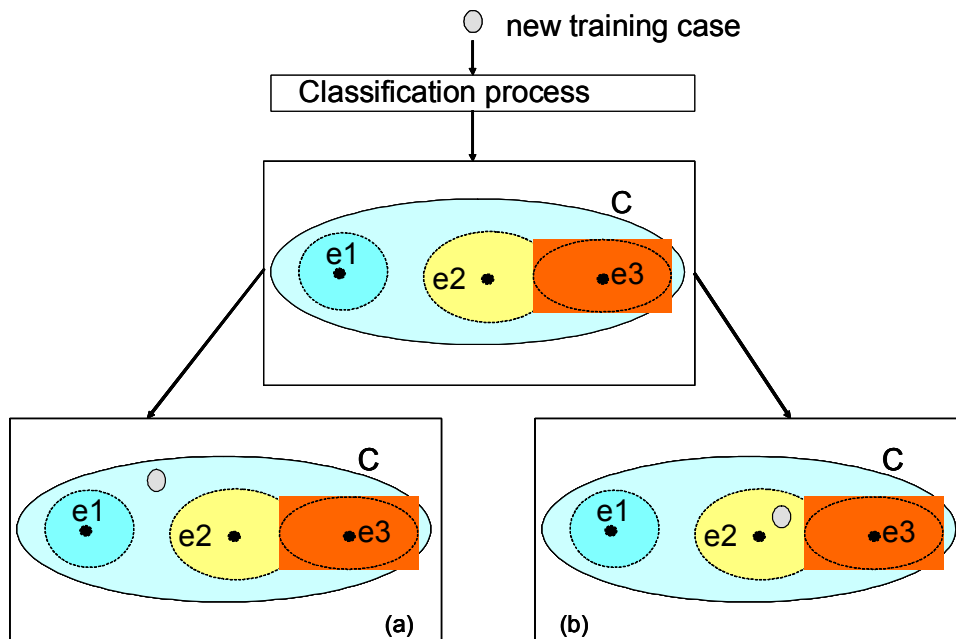


Figure 4: Illustration of the incremental learning procedure

But how do we assess which exemplar is better?

To answer this question, we use the notion of family resemblance that Wittgenstein used when developing his philosophy of language. When Wittgenstein presented his early discussions of the meaning of words, he considered words to have clear meaning and semantics, but as his thoughts matured, he realized that this was inadequate since words such as “games” could not be defined using sufficiency and necessary conditions (Wittgenstein, 1921, 1953). Instead, Wittgenstein postulated that words were characterized by categories of example uses and that the categories could be based on the principle of family resemblance. More recently, Addis et al. (2004), have also argued that this kind of paradigm shift is necessary if we are to go beyond the capabilities of current systems and develop socially sensitive systems.

This idea is explored further by several authors, most notably by Rosch and Mervis (1975) who suggest that a good exemplar is one that satisfies the following properties:

- *Focality*: That is, it has high family resemblance with those it represents
- *Peripherality*: That is, it has low resemblance with those considered outside the family.

Given our representation, we can interpret family resemblance as the probability of an exemplar representing a point, which in turn can be used to compute focality and peripherality. The difference between the focality and the peripherality can then be used as a measure of the prototypicality of the exemplar.

3. Empirical Evaluation on Zoo, Votes and Audiology

The model has been implemented and tested on the Zoo, Votes and Audiology data sets available from the UCI Machine Learning repository (Blake and Merz, 1998). The experiments utilised a 70/30 training/testing split and were repeated with 20 random trials. The following subsections summarise the results

3.1 Results on the Zoo data

Figure 5 shows the exemplars obtained for three of the classes together with the number of cases, accuracy obtained and features. A ‘Y’ marks the presence of a feature and a blank represents the absence of a feature. These exemplars are not only accurate but are in line with our intuition about these classes.

	Class 1		Class 2	Class 6
Cases	31		20	8
Accuracy	98%		99%	100%
Feature	dolphin	cheetah	lark	housefly
hair		Y		Y
feathers			Y	
eggs			Y	
milk	Y	Y		
airborne			Y	Y
aquatic	Y			
predator	Y	Y		
toothed	Y	Y		
backbone	Y	Y	Y	
breathes	Y			
venomous				
fins	Y			
legs		4	2	6
tail	Y	Y	Y	Y
domestic				
catsize	Y	Y		

Figure 5: Some exemplars for three of the Zoo classes

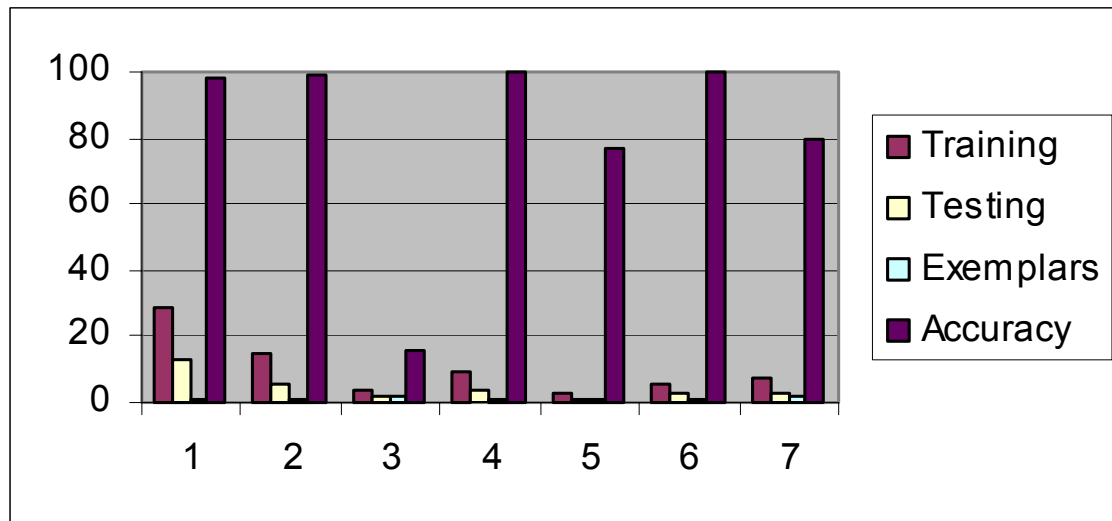


Figure 6: Overall results for the Zoo data set

Figure 6 shows the overall results in terms of the accuracy, no of exemplars, the testing and training cases. The accuracy for the classes 1,2,4,5,6 and 7 is very good. The accuracy for class 3 is poor. It could be argued that this is not surprising given that it only has a few training examples.

However, class 5 also has only a few training examples. So what is the difference? Figure 7, below, shows the examples available for the two classes. From this is clear that class 3 is much more diverse than class 5, explaining the difference in performance. In general, the model developed needs more exemplars and training data for regions that are more diverse, which is consistent with our intuition.

	Class 3 16%						Class 5 77%			
	pitviper	seasnake	slowworm	tortoise	tuatara		frogA	frogB	newt	toad
hair										
feathers										
eggs	Y		Y	Y	Y		Y	Y	Y	Y
milk										
airborne										
aquatic		Y					Y	Y	Y	Y
predator	Y	Y	Y		Y		Y	Y	Y	
toothed	Y	Y	Y		Y		Y	Y	Y	Y
backbone	Y	Y	Y	Y	Y		Y	Y	Y	Y
breathes	Y		Y	Y	Y		Y	Y	Y	Y
venomous	Y	Y						Y		
fins										
legs				4	4		4	4	4	4
tail	Y	Y	Y	Y	Y				Y	
domestic										
catsize				Y						

Figure 7: Class 3 vs Class 5

3.2 Results on the Votes data

The votes data set records the voting behaviour of USA congressmen on 16 issues and their party affiliation, classified into Democrats or Republicans.

On average, 2 exemplars were retained to achieve an accuracy of 96% for the republicans category and 4 exemplars were retained to achieve an accuracy of 84% for the democrats category. Thus the model works very well for this data set. An interesting issue that we investigated on this data set was: Does increasing the number of exemplars retained improve performance? The results obtained, again with 20 repeated random trials for each case, are presented in Figure 8.

The results suggest this is not necessarily the case and indeed, an effect equivalent to overtraining in neural networks or decision tree learning appears to occur.

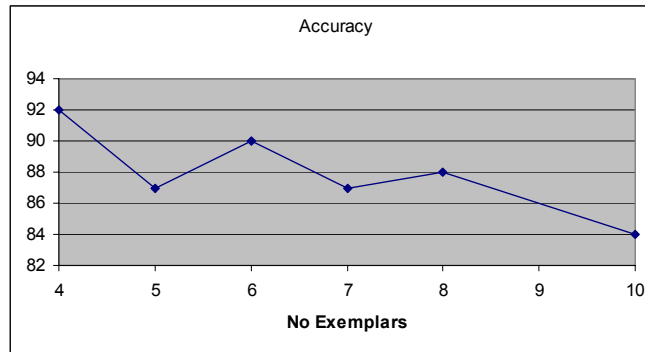
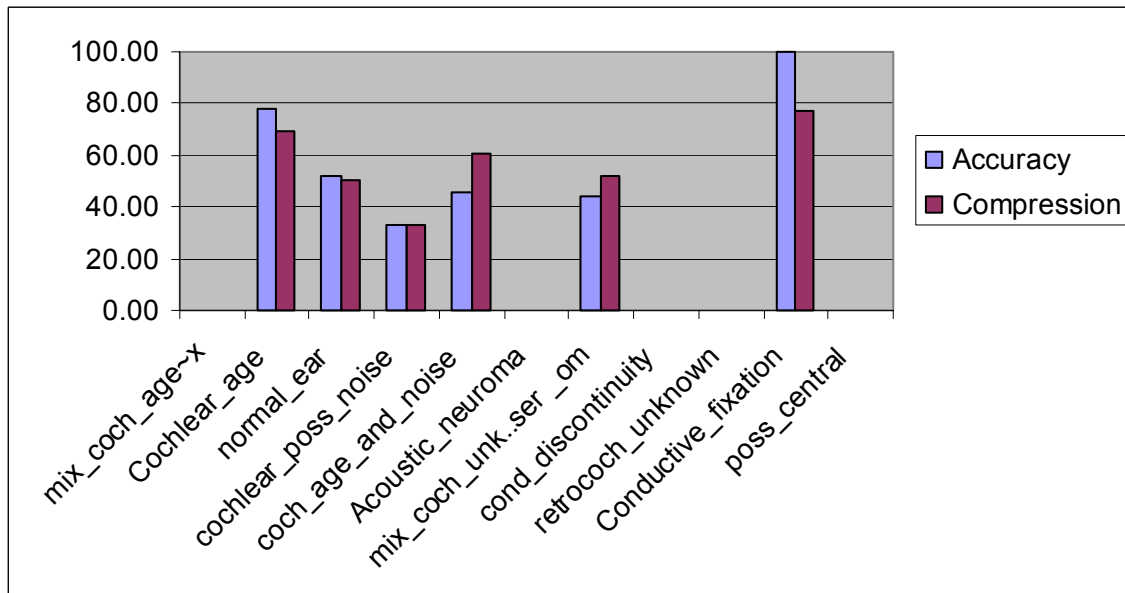


Figure 8: Effect of increasing the number of exemplars on Votes

3.3 Results on the Audiology data

The Audiology data set is the most challenging, since it has noisy data, incomplete information and has some inconsistencies.



The results on the audiology data set are much more varied, but close to those achieved by Bareiss (1989) when the exemplars were hand crafted with the aid of experts for his PROTOS system³. A general characteristic of the results is that forcing an increase in the number of exemplars retained reduces the accuracy and a good indication of the accuracy within a category is the proportion of training cases the exemplars represent. To conclude, these results show some promise in the notion that data can be characterised by a few exemplars which can be learned by machine.

4. Conclusion

Much of what we do appears to rely on experience. If this is true, then we must somehow be able to learn and utilize exemplars that represent cases. This raises interesting questions about whether it is possible to develop a computational system capable of learning exemplars and achieving reasonable performance.

The research presented in this paper tackles this problem and has developed a model for automatically identifying exemplars that has foundations in Bayesian networks and uses the principle of family resemblance.

The model has been implemented and an empirical evaluation shows promising results. The idea that a small number of exemplars can represent a large amount of data is a little radical, though appealing given the vast amounts of data we are amassing. The empirical results, that result in reasonable accuracy with significantly fewer cases, in itself suggest that this idea has practical value.

To conclude, given the initial results, several philosophical questions, of the kind that were considered some years ago (Anderson (1990); Ahn and Medin (1992)) are re-opened, but with some new evidence that it is possible to obtain exemplars computationally: What is the relationship between exemplars and categorization and the kind of exemplars learned by our model? How are exemplars learned by us? Is novice vs expert behaviour distinguished by the extent to which optimal exemplars are developed? We hope these questions are revisited in the light of what has been demonstrated as computationally feasible.

References

- Addis, T, Billinge, D., and Visscher, B., 2004, Socially sensitive computing: A necessary paradigm shift for computer science, In Proc. Grand Challenges in Computing, UK CPHC, Newcastle, pp1-6.
- Addis, T. et al., 2004, Tracking irrational sets, Proceeding of Model Based Reasoning in Science and Engineering, University of Pavia, Italy.
- Ahn, W., and Medin, L., 1992, A two-stage model of category construction, *Cognitive Science*, 16, pp 81-121.
- Anderson, J.R., 1990, *The Adaptive Character of Thought*, Erlbaum, Hillsdale, NJ.
- Bareiss, R. 1989, *Exemplar-based Knowledge Acquisition. A Unified Approach to Concept Representation, Classification, and Learning*, Academic Press Inc. San Diego, CA, U.S.A.
- Blake, C. and Merz, C., 1998, *UCI Repository of Machine Learning Databases*, [<http://www.ics.uci.edu/mllearn/MLRepository.html>], Irvine, CA: University of California, U.S.A.

Pearl, J., 1991, Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, Palo Alto, CA, U.S.A.

Rodríguez, A and Vadera, S., 1999, PEBM: A probabilistic exemplar based model, Proceeding of the International Joint Conference on AI, pp242-247.

Rosch, E. and Mervis, C.B., 1975, Family resemblance studies in the internal structure of categories, *Cognitive Psychology* 7, pp 573—605.

Wittgenstein, L. 1921, *Tractus Logico-Philosophicus*, English edition, 1961, Routledge and Kegan Paul, London.

Wittgenstein, L. 1953, *Philosophical Investigations*, Blackwell, Oxford, UK.